STAT C206B: Topics in Stochastic Processes

Feynman Liang*
Department of Statistics, UC Berkeley

Last updated: February 25, 2020

Contents

1 Lecture 10

2020-02-20

Let $\xi \in S$ and $\eta \in T$, $\xi \overset{d}{=} \tilde{\xi}$.

μ the regular conditional distribution of η given ξ.

$(\xi, \eta) \overset{d}{=} (\tilde{\xi}, \tilde{\eta})$.

Theorem 1 (*Fubini/Tonelli*)

Corollary 2 (*extended Minkowski inequality*)

Let μ, ν, and f be such as in the previous theorem ?? and assume $\mu f(t) = \int f(s,t)\mu(ds)$ exists for $t \in T$. Write $\|f\|_p(s) = (\nu \mid f(s,\cdot)|^p)^{1/p}$. Then

$$\|\mu f\|_p \leq \mu \|f\|_p, \quad p \geq 1$$

Proof. Follows from Hölder's inequality in the usual way. ☐

Recall the regular Minkowski inequality: for $g_k : T \to \mathbb{R}$ and $\alpha_k \geq 0$ for $k \in [n]$,

$$\|\sum_{i=1}^n \alpha_i g_i\|_p \leq \sum_{i=1}^n \alpha_i \|g_i\|_p$$

This is a special case of the corollary by taking $S = [n]$, $\mu(\{k\}) = \alpha_k$, and $f(k,t) = g_k(t)$.

1.1 Ergodic theory

{sub:Ergodic ␣theory}

Fix measure space $(S, \mathcal{S}, \mu)$

Definition 3

A transformation T on S is *μ-preserving* or *measure-preserving* if $\mu \circ T^{-1} = \mu$.
Equivalently, if ξ is a random element of S with distribution μ, T is measure-preserving iff $T\xi \overset{d}{=} \xi$.

Example 4

For a random sequence $\xi = (\xi_i)_i$, let θ denote the left shift operator $\theta(\xi_i)_i = (\xi_{i+1})_i$. ξ is *stationary* if

*feynman@berkeley.edu

$\theta\xi \stackrel{d}{=} \xi$, i.e. θ is measure-preserving for the distribution of ξ.

Lemma 5 (*Statinarity and invariance*)

For random $\xi \in S$ and measurable T on S, $T\xi \stackrel{d}{=} \xi$ iff $(T^n\xi)$ is stationary, in which case $(f \circ T^n\xi)$ is stationary for every measurable f.
Conversely, any stationary random sequence admits such a representation.

Proof. Assuming $T\xi \stackrel{d}{=} \xi$,

$$\theta(f \circ T^n\xi) = (f \circ T^{n+1}\xi) = (f \circ T^nT\xi) \stackrel{d}{=} (f \circ T^n\xi)$$

Conversely, if $\eta = (\eta_i)$ is stationary then $\eta_n = \pi_0(\theta^n\eta)$ with $\pi_0(x_0, x_1, \ldots) = x_0$. $\theta\eta \stackrel{d}{=} \eta$ by stationarity of η. □

Let $\mathcal{S}^\mu$ denote the μ-completion of $\mathcal{S}$.

Definition 6

A set $I \subset S$ is *invariant* if $T^{-1}I = I$ and *almost invariant* if $T^{-1}I = I$ μ-a.e. in the sense that $\mu(T^{-1}I\Delta I) = 0$.
Since inverse images preserve set operations, the classes $\mathcal{I}$ and $\mathcal{I}'$ of invariant sets in $\mathcal{S}$ and almost invariant sets in $\mathcal{S}^\mu$ form σ-fields in S, called the *invariant* and *almost invariant* σ-fields.
f measurable function on S is *invariant* if $f \circ T \equiv f$ and *almost invariant* if $f \circ T = f$ μ-a.e..

Example 7

$S = \{0,1\}^\infty$, $T = \theta$, then $I = \{(x_i)_i : x_k = 0 \text{ i.o. }\}$ is an invariant set.

We are heading to the ergodic theorem; if $\xi = (\xi_i)_i$ is stationary S-valued sequence, $f : S \to \mathbb{R}_+$ $\mathcal{S}$-measurable, then $\frac{1}{denom}\sum_{0\le k\le n} f(\xi_k)$ converges almost surely. This is a generalization of the law of large numbers, where this limit converges to the constant $f(\xi_0)$.

In general, the limit may be random rather than a constant.

Example 8

Let ξ have distribution $\frac{1}{2}\delta_{(0,0,\ldots)} + \frac{1}{2}\delta_{(1,1,\ldots)}$. Then the limit is random.

Lemma 9

For measure μ and measurable transform T on S, let $f : S \to S'$ be measurable mapping into Borel space S'.
Then f is invariant or almost invariant iff it is $\mathcal{I}$-measurable or $\mathcal{I}'$-measurable respectively.

Proof. TODO □

Write $\mathcal{I}^\mu$ for the μ-completion of $\mathcal{I}$ in $\mathcal{S}^\mu$, the σ-field generated by $\mathcal{I}$ and μ-null sets in $\mathcal{S}^\mu$.

Theorem 10

For any distribution μ and μ-preserving transform T on S, the associated invariant and almost invariant σ-fields $\mathcal{I}$ and $\mathcal{I}'$ respectively are related by $\mathcal{I}' = \mathcal{I}^\mu$.

$\mu(A\Delta B)$ is a pseudometric on sets.

Proof. If $J \in \mathcal{I}^\mu$, then $\mu(I\Delta J) = 0$ for some $I \in \mathcal{I}$. Since T is μ-preserving, we get ??? □

Definition 11

T is *ergodic* for μ or simply μ-*ergodic* if the invariant σ-field $\mathcal{I}$ is μ-trivial in the sense $\mu I \in \{0,1\}$ for every $I \in \mathcal{I}$.
Depending on viewpoint, we may also say μ is ergodic for T.

Randomm ξ is ergodic if $\Pr[\xi \in I] \in \{0,1\}$ for any $I \in \mathcal{I}$, i.e. $\mathcal{I}_\xi = \xi^{-1} I$ in Ω is Pr-trivial.

Lemma 12

Let $\xi \in S$ be random with distribution μ, T a μ-preserving mapping on S. Then ξ is T-ergodic iff $(T^n \xi)$ is θ-ergodic in which case $(f \circ T^n \xi)$ is θ-ergodic for every measurable f.

Write $\mathcal{I}_\xi = \xi^{-1}\mathcal{I}$.

Theorem 13 (*Ergodic theorem, Birkhoff*)

Let ξ be a random element in S with distribution μ, T a μ-preserving map on S with invariant σ-field $\mathcal{I}$. Then for any measurable $f \geq 0$ on S

$$n^{-1} \sum_{0 \leq k < n} f(T^k \xi) \xrightarrow{as} \mathbb{E}[f(\xi) \mid \mathcal{I}_\xi]$$

The same convergence holds in L^p for some $p \geq 1$ when $f \in L^p(\mu)$.

The LHS is a limit of Cesaro means.

Definition 14

Distributions μ_n on $\mathbb{N}$ are *asymptotically invariant* if $\|\mu_n - \mu_n * \delta_s\| \to 0$ for every $s \in \mathbb{N}$ where $\|\cdot\|$ denotes the total variation norm, i.e.

$$\|\mu_n - \mu_n * \delta_s\| = \sum_k |\mu_n(\{k\}) - \mu_n * \delta_s(\{k\})|$$

and $*$ denotes convolution, i.e.

$$\mu_n * \delta(\{k\{) = \begin{cases} \mu_n(k-s), & \text{if } k \geq s \\ 0, & \text{ow} \end{cases}$$

Example 15

$\mu_n = n^{-1} \sum_{0 \leq k < n} \delta_k$.

Corollary 16 (*Extended mean ergodic theorem*)

For any $p \geq 1$, consider on $\mathbb{N}$ a stationary L^p-valued process X and some asymptotically invariant μ. Then

$$\mu_n X = \int_{\mathbb{N}} X_s \mu_n(ds) = \sum_{kin\mathbb{N}} X_k \mu_n(\{k\}) \to \bar{X} := \mathbb{E}[X \mid \mathcal{I}_X]$$

in L^p.

Proof. Let $\nu_m = m^{-1} \sum_{0 \leq k < m} \delta_k$ so $\nu_m X \to \bar{X}$ in L^p by mean ergodic theorem. By generalized Minkowski's inequality along with stationarity of X, invariance of $\bar{X}$ and dominated convergence, we get as $n \to \infty$ and then $m \to \infty$

$$\begin{aligned}
\|\mu_n X - \bar{X}\|_p &\leq \|\mu_n X - (\mu_n * \nu_m) X\|_p + \|(\mu_n * \nu_m) X - \bar{X}\|_p \\
&\leq \|\mu_n - \mu_n * \nu_m\| \|X_0\|_p + \int \|(\delta_s * \nu_m) X - \bar{X}\|_p \mu_n(ds) \\
&\leq \|X\|_p \int \|\mu_n - \mu_n * \delta_t\| \nu_m(dt) + \|\nu_m X - \bar{X}\|_p \to 0
\end{aligned}$$

□

2 Lecture 11

2020-02-25

Definition 17

A finite or infinite sequence $(\xi)_{i\in\mathbb{N}}$ in measurable space $(S, \mathcal{S})$ is *exchangeable* if $(\xi_{k_i})_{i=1}^m \stackrel{d}{=} (\xi_i)_{i=1}^m$ for any $(k_i)_{i=1}^m$ distinct.
ξ is *contractible* if the above holds for increasing $k_1 < \cdots < k_m$.

Definition 18

(ξ_i) is *conditionally iid* if $\Pr[\xi \in \cdot \mid \mathcal{F}] \stackrel{as}{=} \nu^\infty$ for some σ-field $\mathcal{F}$ and random probability measure ν on S.

We can view ν as a probability kernel from the ambient probability space $(\Omega, \mathcal{A})$ to S, or equivalently a random element of the space $\mathcal{M}_1(S)$ of probability measures on S equipped with σ-field generated by projection maps $\pi_B : \mu \mapsto \mu B$ for $B \in \mathcal{S}$. Notice ν is almost surely $\mathcal{F}$-measurable, so we can take $\mathcal{F} = \sigma(\nu)$. Taking expected values shows that ξ is *mixed iid* i.e.

$$\Pr[\xi \in \cdot] = \mathbb{E}\nu^\infty = \int_{\mathcal{M}_1(S)} m^\infty \Pr[\nu \in dm]$$

Definition 19

A measurable space $(S, \mathcal{S})$ is a *Borel space* if there exists a measurable bijection $f : S \to B$ for some Borel subset $B \in \mathcal{B}(\mathbb{R})$ such that f^{-1} is measurable.

Theorem 20

For (ξ) an infinite sequence of random elements in measurable Borel space S, contractible = exchangeable = conditionally iid.

Borel is essential

S Borel is essential, Blackwell has a counterexample.

Proof. Suffices to show contractible implies conditionally iid, since the other implications are obvious.

Let $\mathcal{I}_\xi = \xi^{-1}\mathcal{I}$ where $\mathcal{I}$ denotes the shift-invariant σ-field in $(S, \mathcal{S})^\infty$. The conditional distribution $\nu = \Pr[\xi_1 \in \cdot \mid \mathcal{I}_\xi]$ exists since S is Borel (hence regular conditional distributions exist).

Fix $I \in \mathcal{I}$ and bounded measurable functions $\{f_i\}_{i=1}^m$ on S. Since shift invariance means $\{\xi \in I\} = \{\theta_k\xi \in I\}$ for all k, by contractibility, extended mean ergodic theorem, and dominated convergence theorem, as $n \to \infty$

$$\begin{aligned}
\mathbb{E}\mathbb{1}_I(\xi) \prod_{k\le m} f_k(\xi_k) &= n^{-m} \sum_{j_1,\dots,j_m \le n} \mathbb{E}\mathbb{1}_I(\xi) \prod_{k\le m} f_k(\xi_{kn+j_k}) \\
&= \mathbb{E}\mathbb{1}_I(\xi) \prod_{k\le m} n^{-1} \sum_{j\le n} f_k(\xi_{kn+j_k}) \\
&\to \mathbb{E}\mathbb{1}_I(\xi) \prod_{k\le m} \nu f_k
\end{aligned}$$

The first equality reindexes $kn + j_k \in [kn+1, (k+1)n]$ using contractibility and then averages over all n^m choices of $(j_k)_{k=1}^m \in n^m$. The second equality simply rearranges sum of products into product of sums, which look like Cesaro sums except they start from kn rather than 1. Nevertheless, for fixed $k \in [m]$ the sums

$$n^{-1} \sum_{j\le n} f_k(\xi_{kn+j_k}) \to_{L^p(\Pr)} \mathbb{E}[f_k(\xi_1) \mid \mathcal{I}_\xi] = \nu f_k$$

for any p by the extended mean ergodic theorem, so in particular convergence in probability holds. Since it is bounded, by dominated convergence we may move the limit inside the expectation.

But since there are no ns on the initial and final expression, we have in fact

$$\mathbb{E}\mathbb{1}_I(\xi) \prod_{k\le m} f_k(\xi_k) = \mathbb{E}\mathbb{1}_I(\xi) \prod_{k\le m} \nu f_k$$

The functions $(s_i)_1^m \mapsto \prod_{k\le m} f_k(\xi_k)$ form a π-system, so by a monotone-class argument and the definition of conditional probabilities

$$\Pr[\xi \in B \mid \mathcal{I}_\xi] \overset{as}{=} \nu^\infty B$$

for all $B \in \mathcal{S}^\infty$. This is the statement of conditionally iid with $\mathcal{F} = \mathcal{I}_\xi$. □

Proposition 21 (*Conditional independence, Doob*)

{prop:cond-indep-doob}

For σ-fields $\mathcal{F}$, $\mathcal{G}$, and $\mathcal{H}$, $\mathcal{F} \perp_\mathcal{G} \mathcal{H}$ iff

$$\Pr[H \mid \mathcal{F}, \mathcal{G}] \overset{as}{=} \Pr[H \mid \mathcal{G}]$$

for all $H \in \mathcal{H}$.

Lemma 22 (*Contraction and independence*)

If random elements $(\xi, \eta) \overset{d}{=} (\xi, \zeta)$ and $\sigma(\eta) \subset \sigma(\zeta)$ (at least as much information in ζ as there is in η) then $\xi \perp_\eta \zeta$ (ζ is built from η plus something additional which is conditionally independent given η).

Proof. Fix measurable B in range of ξ, define $\mu_1 = \Pr[\xi \in B \mid \eta]$ and $\mu_2 = \Pr[\xi \in B \mid \zeta]$.

(Useful facts about martingales: if $(\mathcal{F}_n)$ is a filtration and Z is a random variable, then $X_n = \mathbb{E}[Z \mid \mathcal{F}_n]$ is a martingale. Also, L^2 martingales have orthogonal increments.) Then (μ_1, μ_2) is a bounded martingale with $\mu_1 \overset{d}{=} \mu_2$ (so in particular $\mathbb{E}\mu_2^2 = \mathbb{E}\mu_1^2$), hence $\mathbb{E}(\mu_2 - \mu_1)^2 = \mathbb{E}\mu_2^2 - \mathbb{E}\mu_1^2 = 0$ which implies $\mu_1 \overset{as}{=} \mu_2$ □

Second proof of Theorem contractible=exchangeable=conditionally iid. If ξ is contractible, then

$$(\xi_m, \theta_m\xi) \overset{d}{=} (\xi_k, \theta_m\xi) \overset{d}{=} (\xi_k, \theta_n\xi) \qquad k \le m \le n$$

Let $\mathcal{T}_\xi = \cap_n \sigma(\theta_n\xi)$ be the tail σ-field and fix $B \in \mathcal{S}$. By the previous lemma and reverse martingale convergence, as $n \to \infty$

$$\Pr[\xi_m \in B \mid \theta_m\xi] = \Pr[\xi_k \in B \mid \theta_m\xi] = \Pr[\xi_k \in B \mid \theta_n\xi] \overset{as}{\to} \Pr[\xi_k \in B \mid \mathcal{T}_\xi]$$

where we used the lemma to assert that conditioning on less information $\theta_n\xi$ compared to $\theta_m\xi$ ($n > m$) doesn't change things, and as n varies we have a reverse martingale.

Since the two sides dont have n appearing we in fact have

$$\Pr[\xi_m \in B \mid \theta_m\xi] \overset{as}{=} \Pr[\xi_m \in B \mid \mathcal{T}_\xi] \overset{as}{=} \Pr[\xi_1 \in B \mid \mathcal{T}_\xi]$$

Here, the first relation yields $\xi_m \perp_{\mathcal{T}_\xi} \theta_m\xi$ for all $m \in \mathbb{N}$, so iterating gives $(\xi_i)_i$ are conditionally independent given $\mathcal{T}_\xi$. The second relation shows the conditional distributoins agree almost surely, so together we have that shown conditionally iid with $\mathcal{F} = \mathcal{T}_\xi$ and $\nu = \Pr[\xi_1 \in \cdot \mid \mathcal{T}_\xi]$. □

Definition 23

A contractible/exchangeable sequence ξ is *extreme* if its distirbution μ cannot be expressed as a nontrivial mixture $p\mu_1 + (1-p)\mu_2$ of exchangeable/contratable distributions.
We use $\mathcal{F} \overset{as}{=} \mathcal{G}$ (σ-fields agree almost surely) to mean that the Pr-completions agree.

Proposition 24 (*Uniqueness and extremality*)

For ξ an infinite exchangeable sequence in a Borel space $(S, \mathcal{S})$ such that $\Pr[\xi \in \cdot \mid \mathcal{F}] \overset{as}{=} \nu^\infty$ for some σ-field $\mathcal{F}$ and random probability measure ν on S,

1. *ν is a.s. unique, ξ-measurable, and given by limiting empirical probability distribution*

$$n^{-1} \sum_{k \le n} \mathbb{1}_B(\xi_k) \overset{as}{\to} \nu B \qquad B \in \mathcal{S}$$

2. *$\mathcal{F} \perp_\nu \xi$ and $\mathcal{F} \subset \sigma(|xi)$ implies $\mathcal{F} \overset{as}{=} \sigma(\nu)$*
3. *$\mathcal{L}(\xi) = \int m^\infty \mu(dm)$ (the law of ξ is a mixture of infinite products) iff $\mu = \mathcal{L}(\nu)$.*
4. *$\mathcal{L}(\xi)$ is extreme iff ν is a.s. non-random.*

As a result of uniqueness, we will refer to the random ν as the *directing random measure* of ξ and say ξ is directed by ν.

Proof. (1) Fix measurable $f \geq 0$ on S. By disintegration theorem (??, taking $\nu = \nu^\infty, \eta = \nu f$) and SLLN

$$\Pr\left\{n^{-1}\sum_{k\leq n} f(\xi_k) \to \nu f\right\} = \mathbb{E}\nu^\infty\left\{x; n^{-1}\sum_{k\leq n} f(x_k) \to \nu f\right\} = 1$$

This proves convergence and a.s. uniqueness of ν follows by a monotone class argument with a countable class of nonnegative functions that generate $\mathcal{S}$ (because we are dealing with a Borel space).

(2) is clear from Proposition 21 and disintegration corollary ????

(3) Let $\tilde{\nu}$ be a random probability measure with distribution μ on $\mathcal{M}_1(S)$. By extension lemma ??? we can construct a random $\eta = (\eta_j)$ on S satisfying $\Pr[\nu \in \cdot \mid \tilde{\nu}] \overset{as}{=} \tilde{\nu}^\infty$. Then

$$\Pr \circ \eta^{-1} = \mathbb{E}\tilde{\nu}^\infty = \int m^\infty \mu(dm)$$

and comparing with definition of conditionally iid we see $\nu \overset{d}{=} \tilde{\nu}$ implies $\xi \overset{d}{=} \eta$.

Conversely, if $\xi \overset{d}{=} \eta$ then applying (1) to both ξ and η shows $\nu f \overset{d}{=} \tilde{\nu} f$ for all measurable $f \geq 0$ on S, hence $\tilde{\nu} \overset{d}{=} \nu$ again by a monotone class argument.

(4) Use (3) and the fact that a probability measure on $\mathcal{M}_1(S)$ is extreme iff it is degenerate (concentrated at a single point). $\square$